






Natural Fracture Network Model Using Machine Learning Approach

Timur Merembayev  and Yerlan Amanbek  

Department of Mathematics, School of Sciences and Humanities,
Nazarbayev University, Kabanbay batyr 53, Astana, Kazakhstan
yerlan.amanbek@nu.edu.kz

Abstract. A fracture network model is a powerful tool for characterizing fractured rock systems. In this paper, we present the fracture network model by integrating a machine learning algorithm in two-dimensional setting to predict the natural fracture topology in porous media. We also use a machine learning algorithm to predict the fracture azimuth angle for the natural fault data from Kazakhstan. The results indicate that the fracture network model with LightGBM performs better in designing a fracture network parameter for hidden areas based on data from the known area. In addition, the numerical result of the machine learning algorithm shows a good result for randomly selected data of the fracture azimuth.

Keywords: Fracture characterization · Machine learning · LightGBM

1 Introduction

The mapping of natural fracture networks plays a significant role in predicting the hydrocarbon production of reservoir, flow and transport problems in oil and gas production. Due to insufficient data on the subsurface characteristic, predicting the fracture network is still a challenge.

The current literature contains numerous publications that explore the fracture model in the subsurface by utilizing geomechanical conditions that necessitate significant computational power. In recent years, several publications have been published documenting the fracture model in porous media. These works focus on analyzing the fracture behavior in the coupled system and give insights into complex mechanics [1, 2, 19, 24]. However, the complexity of geomechanical models can lead to challenges in verifying simulation results because the model's inputs, including physical properties, boundary conditions, and others, exhibit significant uncertainty and spatial variations.

Considering the problem of mapping fracture network from a geomechanical point of view, the direction of fractures depends on the properties of the subsoil, such as geomechanical features, intersections of fractures, and others [6]. The stress field near fracture nodes is critical in modeling fracture propagation. The research [19, 24] showed that energy minimization is important in systematically

© The Author(s)

O. Gervasi et al. (Eds.): ICCSA 2023 Workshops, LNCS 14107, pp. 384–397, 2023.

https://doi.org/10.1007/978-3-031-37114-1_26

modeling the fracture path using the phase field model. Therefore, the influence of adjacent fractures should be considered when modeling the direction of the fracture.

To overcome these challenges, recent research has focused on developing fracture models that can effectively model complex fracture networks. These models aim to provide a faster and more realistic simulation process that can be used in real-time decision making for reservoir management. One of the effective tools for reducing uncertainty in porous media is stochastic simulation algorithms [8]. These methods incorporate geological data of the fracture network, including fracture azimuth distributions, location and length. While the fracture network model may not always generate the exact topology of porous media, it still provides a suitable configuration for the flow and transport process [3, 4, 12]. The application of such tasks can be useful for ground remediation, CO₂ sequestration, hydrocarbon production, and others. Data collected near a wellbore, or preliminary seismic process from a specific location is a good candidate for training geostatistical models to simulate geological configuration for unknown near locations. These stochastic approaches have gained considerable attention from researchers in their ability to reduce uncertainty in geological modeling.

The Fracture Network model is a powerful tool for simulating fluid flow in fractured porous media. The accuracy of the model depends on the ability to represent the geometry of fractures in the subsurface accurately. There are several approaches that can be used to create the geometry of fractures, including geostatistical analysis, numerical modeling, and field observations. A hybrid approach that combines these methods can provide the most accurate representation of the subsurface while minimizing uncertainties.

Machine learning solves the problems of developing learning algorithms, including unsupervised learning, supervised learning, data dimensionality reduction methods, and feature evaluation. Unsupervised learning allows the identification of stable and rare states of systems; regression and classification are used to predict and identify the state of objects. The use of classical or modern machine learning methods is determined by the amount of data at the disposal of researchers [10, 20].

In the literature, some works have shown the application of machine learning methods to reduce uncertainty in geoscience applications [5, 16, 17]. The well log data is also used by machine learning algorithms such as KNN, Decision Tree, Random Forest, XGBoost, and LightGBM in order to detect multilabel lithofacies classification in [18].

The authors of the paper [15] proposed geostatistical modeling of the fracture system using pattern statistics. The images used multiple-point statistics (MPS) to model the fracture network and its propagation. The authors of [9] used the same MPS for simulating a fracture network; the authors trained the model on satellite images to build fractures on the surface and deep into the earth. This approach is closest to machine learning algorithms. MPS is a popular method for constructing a fracture network if fracture images are available. The MPS method is used in studies [7, 13] to model the fracture network system. The MPS

method considers the available data around fractures. It builds statistical histograms of known zones to model a network of fractures for unknown zones while maintaining properties from the known zone. This method has limitations for the problem under consideration, building a network of fractures underground. Tasks contain a significant amount of uncertainty, and obtaining images is difficult. Training the MPS method requires a large number of images for quality training. In this regard, this approach for subsurface construction of a fractured network will not be considered.

In this paper, we extend the fracture network model [3] by incorporating with machine learning algorithm in 2D to predict the parameters of fractures in the porous media. We also used machine learning algorithms to generate the parameters of fracture topology, including azimuth angle and compared them with the proposed fracture network model. The natural fault data from Kazakhstan (North of Balkhash Lake) was used to verify the proposed model.

2 Methodology

We consider the fractures for the algorithm and use the digitized data from the geological fault zone. In the fracture network model, the fracture is a collection of segments, which is the line between nodes. In other words, the fracture is a graph containing nodes and branches(fracture segments). Segment's midpoint is used to define the segment location in a domain. In Fig. 1, nodes are presented in black circles, and the midpoint of segments is in the blue circle. Based on the defined location of segments we calculate the azimuth angle of each of them. The azimuth angle of the segment is measured between the fracture segment direction and the north vector, see Fig. 1.

2.1 Fracture Network Algorithm

We use machine learning approaches for the probabilistic classification of azimuth fractures. The input data is the 2D fracture network, which contains the coordinates of each fracture segment. Based on the coordinates, we calculated the azimuth of the fracture segment and the distance between the closest fracture segments, see Fig. 2.

Generated features from the known region fracture networks describing fracture geometries and positions are used to predict the azimuth of the fracture network in the unknown region. The key goal of the classification model is to predict the 8 classes of azimuth of the fracture segment. Below, detailed information about the classification of azimuth angle is presented. The steps of generation features from the fracture network are presented below, and as an example, it is for one initial fracture:

- The azimuth of the closest 6 fractures segment in the fracture network (6 parameters);
- The coordinates of the closest 6 fractures segment: X coordinates are 6 features, and Y coordinates are 6 features;

- The distance of the closest 6 fractures segment (polar coordinates);
- The azimuth for the closest 6 fractures segment (polar coordinates).

In Fig. 3, there is a flowchart of the proposed steps of the algorithm. We prepared a training and validation dataset from the fracture network. The target of the training model is an azimuth angle. In the algorithm, the prediction of azimuth angle segments in machine learning model proceeds as follows:

1. We define the 6 neighbor fractures for each fracture segment. From each fracture, we got true azimuth fractures for the training model.
2. We calculate two types of azimuth: First, the azimuth angle of the 6 closest fractures segment, and second, the azimuth angles from the initial fracture to the 6 closest fractures segment. Also, we define distances from the initial fracture to the 6 closest fractures.
3. The azimuth fracture angles range from 0 to 360 and are divided into 8 equal sectors of angle 45 centered at the fracture. The angle ranges related to each azimuth class are provided in Table 2.
4. Preparing the dataset for training and validation is performed in two ways:
 - (a) split the dataset randomly;
 - (b) split the dataset by known and hidden areas of interest.
5. Training the LightGBM models and validating the result of the models on the validation dataset.

After training, the ML model can forecast the class azimuth of a fracture network for the location where the model has been trained.

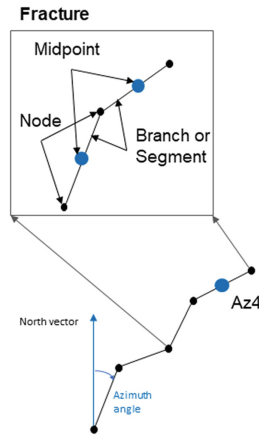


Fig. 1. The fracture network illustrates different node and segment types.

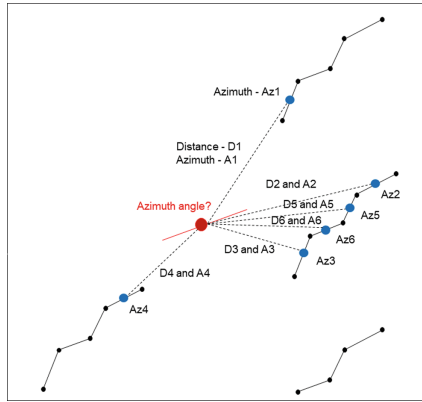


Fig. 2. The searching closest fracture segments.

2.2 Machine Learning Algorithm

Now there are separate machine learning algorithms, each with pros and cons for resolving various problems. In the study, we chose one machine learning algorithm - LightGBM, to verify our hypothesis and classify fracture characterization. The LightGBM algorithm is an open-source framework used for almost tabular data, and our dataset is also tabular. This approach is decision-tree-based, an improved variant of the gradient boosting decision tree (GBDT) algorithm [14].

Another reason is the boosting ensemble algorithms can provide good performance for multi-class imbalanced data. The various problems are multi-class imbalanced data, which have been managed by applying ensemble learning techniques. It is one of the main reasons why we chose LightGBM for this task [23]

The aim of LightGBM algorithm is to obtain an estimate $\widehat{F}(X)$, of the function $F(X)$ mapping X to Y with minimization of the loss function $L(Y, F(X))$.

In gradient boosting, each new b_i algorithm (tree) is added to the already built composition:

$$a_i(x) = a_{i-1}(x) + b_i(x) \tag{1}$$

Such an algorithm corrects the answers of the algorithm $a_i(x)$ to correct answers on the training set. If we consider several algorithms, the algorithm is:

$$a_n(x) = \sum_{i=1}^N b_t(x) \tag{2}$$

For the classification task, the loss function has several options, one of option is:

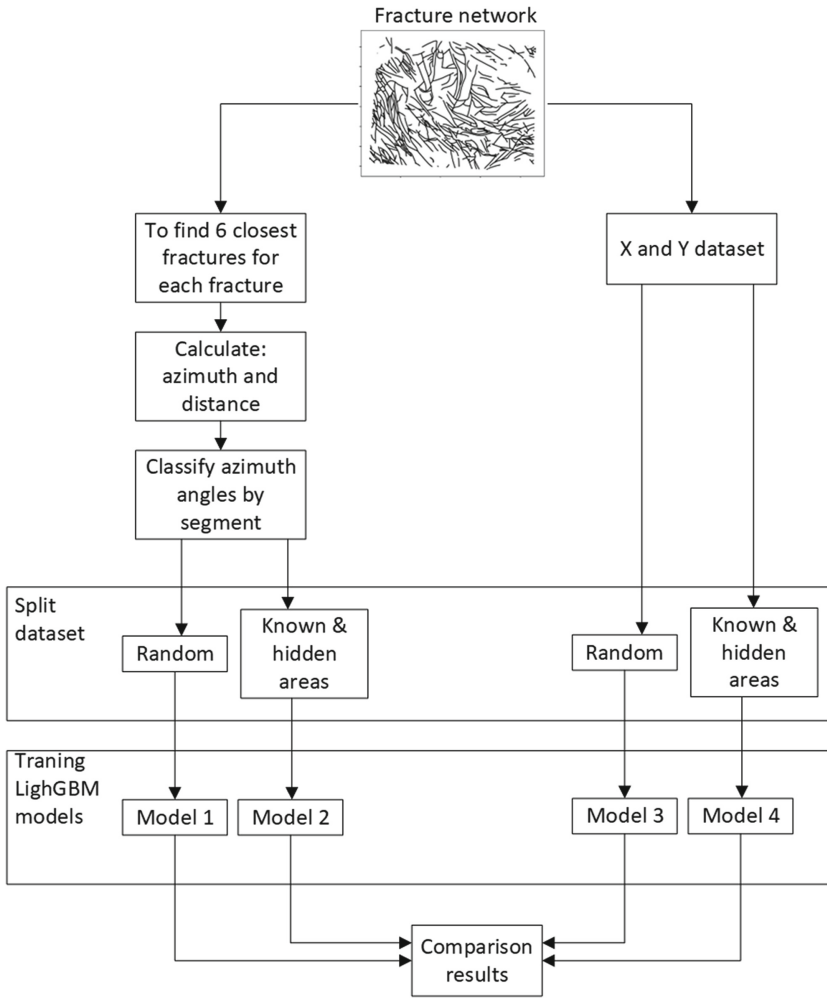


Fig. 3. Flow-chart of the fracture algorithms.

$$L(Y, F(x)) = \log(1 + \exp(-YF(x))) \tag{3}$$

where $F(x) = a_n(x) + s_i$, $s = (s_1, \dots, s_l)$ - vector of shift (correction). Our loss function is:

$$L(Y, F(x)) = \log(1 + \exp(-YF(x))) \tag{4}$$

$$\sum_{i=1}^l \log(1 + \exp(-y_i * (a_n(x_i) + s_i))) \rightarrow \min_s \tag{5}$$

The forecasting performance of the machine learning models is estimated by one statistical indicator - f1 score [11]. The f1 score has a balance between precision and recall. This metric is used when the class distribution has irregular. The f1 metric is a good scoring metric for imbalanced data when a model needs to classify the positives [21].

$$F1 = 2 * \frac{(precision * recall)}{(precision + recall)} \quad (6)$$

2.3 Data Analysis

We train and validate the algorithm in a 31701 km² area north of Balkhash lake, Kazakhstan. The area of interest is near several gold, silver, and copper mines [22]. In addition, there are several actual and possibility mines in the area, see Fig. 4 and Fig. 5. We made digit data of geology and faults for the study area to train and evaluate fracture characterization.

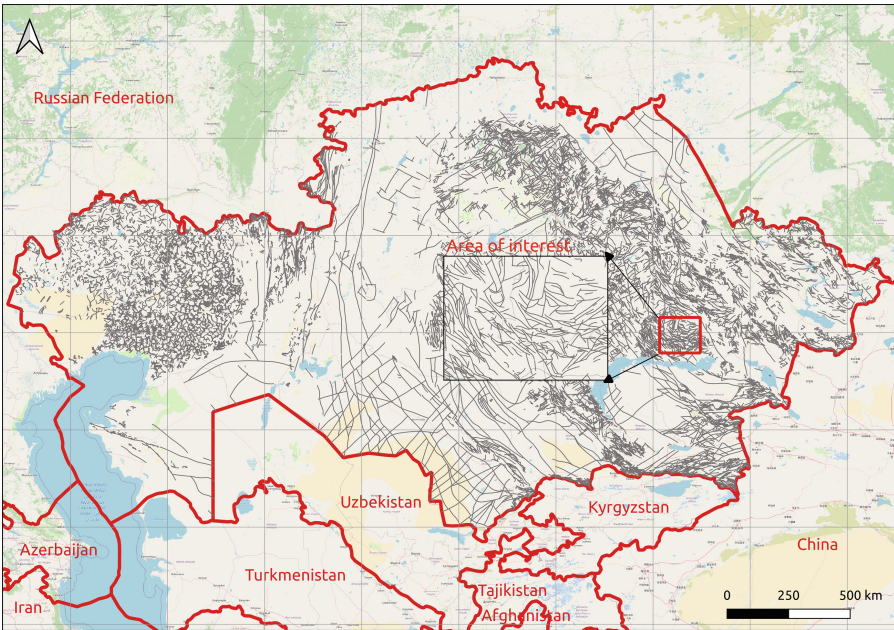


Fig. 4. Area of interest in Kazakhstan [22].

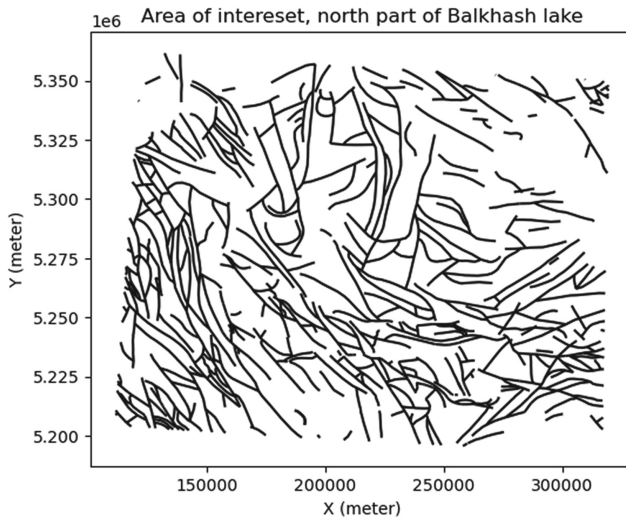


Fig. 5. The geological faults data of Central Kazakhstan.

Figure 6 shows histograms of the azimuth of geology faults. Azimuth angle presented from 0 until 360. The histogram is non-normal distribution because there are several groups of azimuth. Table 1 gives a descriptive statistic of azimuth.

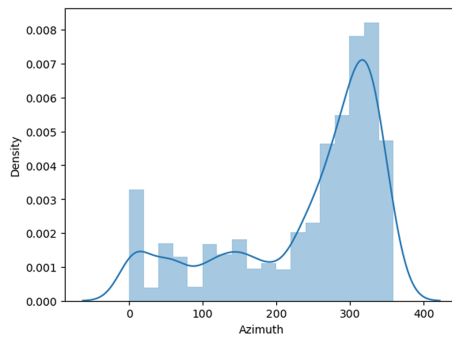


Fig. 6. Histograms of azimuth.

Table 1. The descriptions of azimuth.

| Statistics parameter | Value |
|----------------------|-------|
| Count | 3087 |
| Mean | 241.4 |
| Mode | 351.0 |
| Std | 104.5 |
| Min | 0 |
| 25% | 173.4 |
| 50% | 284.9 |
| 75% | 320.0 |
| Max | 359.0 |

The histogram showed that some segments have a small number of azimuth angles. We classified azimuth by segments. In table 2, azimuth is classified by segments. The 4 segments [0,45), [45,90), [135,180), and [180,225) have not enough account azimuth for training. Therefore we do not use these segments for the training model because a model can not be trained on these segments.

3 Numerical Results

We applied LightGBM to real geological fracture networks to classify the azimuth of fractures. We provide a comparison of the result LightGBM for two cases of splitting dataset and two cases of input dataset:

Table 2. The angle ranges related to each azimuth class for Balkhash dataset.

| Class | Count | Azimuth (Degree) |
|---------|-------|------------------|
| 0 | 7 | [0,45) |
| 0 | 5 | [45,90) |
| class 0 | 122 | [90,135) |
| 0 | 3 | [135,180) |
| 0 | 8 | [180,225) |
| class 1 | 98 | [225,270) |
| class 2 | 231 | [270,315) |
| class 3 | 92 | [315,360) |

- Split training and validation datasets are performed randomly by 80% and 20%, respectively. The dataset contains just locations X and Y; secondly, the input dataset all information about the 6 closest fractures;
- Split training and validation datasets are performed by known and hidden areas, see Fig. 7.

For each prepared dataset, we trained LightGBM models, as mentioned in the flow-chart of the fracture algorithms Fig. 3. Models 1 and 2 used data for 6 closest fractures, with random and known and hidden areas splitting, respectively. Models 3 and 4 used data for location data just X and Y features, with random and known and hidden areas splitting, respectively

3.1 Case with Random Selection

The total dataset for the 6 closest fractures is 535 rows and 30 features. We randomly split the dataset into the training dataset contains 428 rows and 30 features, and the validation dataset contains 117 rows and 30 features. The total dataset for the X and Y fractures is 2961 rows and 2 features. We randomly split the dataset into the training dataset contains 2368 rows and 2 features, and the validation dataset contains 593 rows and 2 features.

The result has been compared on X and Y and by 6 closest fractures datasets by the f1 metrics. The classification report of the models are provided in Table 3 and 4. By considering the score information from the tables, we highlighted that model 3 showed better results for a dataset with X and Y, and every 4 classes have more than 0.64 f1 scores. For the 6 closest fractures, model 1 got less f1 score, but for class 0, it is 0.87, other classes f1 scores are less.

3.2 Case with Known and Hidden Areas

To validate the model for this case, we hide some areas of the fracture network in the center area (cropped from the original domain). In Fig. 7, the red color line is the limit of the crop domain from the original fractures. The black color lines are the original fractures or fractures from the known area.

The total dataset for the 6 closest fractures is 553 rows and 30 features. We took data from a known area containing 461 rows and 30 features, and the validation dataset, the hidden area, contains 92 rows and 30 features. The total

Table 3. Classification report of LightGBM result for X and Y.

| Class | Precision | Recall | F1 | Count |
|-------|-----------|--------|------|-------|
| 0 | 0.63 | 0.64 | 0.64 | 147 |
| 1 | 0.71 | 0.60 | 0.65 | 93 |
| 2 | 0.69 | 0.68 | 0.69 | 165 |
| 3 | 0.69 | 0.74 | 0.71 | 188 |

Table 4. Classification report of LightGBM result for 6 closest fractures.

| Class | Precision | Recall | F1 | Count |
|-------|-----------|--------|------|-------|
| 0 | 0.82 | 0.92 | 0.87 | 25 |
| 1 | 0.46 | 0.35 | 0.40 | 17 |
| 2 | 0.45 | 0.48 | 0.47 | 31 |
| 3 | 0.45 | 0.44 | 0.45 | 34 |

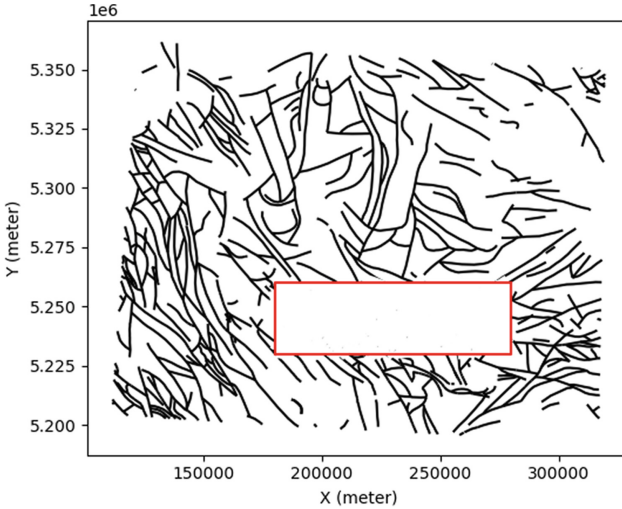


Fig. 7. Hidden zone of fractures from the geological faults data from Kazakhstan.

dataset for the X and Y fractures is 2961 rows and 2 features. We took data from a known area containing 2434 rows and 2 features, and the validation dataset, the hidden area, contains 442 rows and 2 features.

The classification report of the models are provided in Table 5 and 6 for cases with known and hidden areas. The model showed 2 better results for a dataset with 6 closest fractures, each of 4 classes having more than 0.46 f1 scores, and it is better than X and Y datasets (model 4).

Table 5. Classification report of LightGBM result for X and Y.

| Class | Precision | Recall | F1 | Count |
|-------|-----------|--------|------|-------|
| 0 | 0.24 | 0.20 | 0.22 | 107 |
| 1 | 0.30 | 0.26 | 0.28 | 86 |
| 2 | 0.26 | 0.25 | 0.25 | 122 |
| 3 | 0.32 | 0.42 | 0.36 | 127 |

Table 6. Classification report of LightGBM result for 6 closest fractures.

| Class | Precision | Recall | F1 | Count |
|-------|-----------|--------|------|-------|
| 0 | 0.40 | 0.55 | 0.46 | 14 |
| 1 | 0.65 | 0.65 | 0.65 | 20 |
| 2 | 0.59 | 0.80 | 0.62 | 29 |
| 3 | 0.47 | 0.79 | 0.59 | 29 |

4 Discussion

Using the natural fault data from Kazakhstan, we established that a machine learning algorithm could be used for the problem of recreation of a fracture network for a zone with uncertainty. We considered machine learning approaches for the probabilistic classification of azimuth fractures. This approach contains two limitations.

Firstly, machine learning approaches require a lot of data to get a reasonable result. In our case, we have the same problem with an imbalanced dataset, and some classes do not have enough data to train a model. We excluded some classes from the process due to the amount of data that is not reasonable to catch a pattern of fracture in these classes. Therefore, the selected dataset should contain enough data to train and validate a model of machine learning.

Secondly, we concentrated on the fixed length of the fracture segment. The length of the fracture segment defines the fracture propagation from the initial point to the neighbor fracture segment. We set the length of the segment as an average fracture length distribution from a known fracture network. In further research, we will study fracture length, anisotropy, and connectivity available of fractures, it should enable better prediction of fracture network.

5 Conclusion

This paper analyzes the numerical models integrated with LightGBM to classify fracture network azimuth from the Kazakhstan geological data in different scenarios. The findings suggest that the fracture network model with LightGBM shows better results in creating fracture geometry parameters for the unknown area based on known area features. The real fault data from Kazakhstan was applied to different models. The direct model, which uses coordinates with azimuth angles, has a good result in F1 measurement for the randomly selected subset of data.

When comparing the classification results by machine learning algorithm for two datasets with features of only fracture segment coordinates and 6 nearest neighbors, we observed that the model 3 has a good result for the dataset with coordinates in randomly splitting the dataset for training and validation dataset. In the case of hidden zone problem, model 2 predicts better for a dataset containing features with 6 neighbors. This suggests that model 2 captures the key

knowledge of fault patterns in the known zone and applies it to the hidden zone successfully.

In our further research, we intend to concentrate on the regression of azimuth fracture; also, we will apply deep learning algorithms such as LSTM to predict azimuth.

Acknowledgement. The authors wish to acknowledge the support of the research grant, no. AP19575428, from the Ministry of Science and Higher Education of the Republic of Kazakhstan. Authors gratefully acknowledge the support of the Nazarbayev University Faculty Development Competitive Research Grant (NUFDCRG), Grant No. 20122022FD4141.

References

1. Almani, T., Kumar, K.: Convergence of single rate and multirate undrained split iterative schemes for a fractured biot model. *Comput. Geosci.* 1–20 (2022). <https://doi.org/10.1007/s10596-021-10119-1>
2. Almani, T., Lee, S., Wheeler, M.F., Wick, T.: Multirate coupling for flow and geomechanics applied to hydraulic fracturing using an adaptive phase-field technique. In: *SPE Reservoir Simulation Conference*. OnePetro (2017)
3. Amanbek, Y., Merembayev, T., Srinivasan, S.: Framework of fracture network modeling using conditioned data with sequential gaussian simulation. *Arab. J. Geosci.* **16**(3), 219 (2023)
4. Andrianov, N., Nick, H.M.: Modeling of waterflood efficiency using outcrop-based fractured models. *J. Petrol. Sci. Eng.* **183**, 106350 (2019)
5. Andrianov, N., Nick, H.M.: Machine learning of dual porosity model closures from discrete fracture simulations. *Adv. Water Resour.* **147**, 103810 (2021)
6. Berrone, S., Della Santa, F., Pieraccini, S., Vaccarino, F.: Machine learning for flux regression in discrete fracture networks. *GEM - Int. J. Geomath.* **12**(1), 1–33 (2021). <https://doi.org/10.1007/s13137-021-00176-0>
7. Chandna, A., Srinivasan, S.: Modeling natural fracture networks using improved geostatistical inferences. *Energy Proc.* **158**, 6073–6078 (2019)
8. Chandna, A., Srinivasan, S.: Probabilistic integration of geomechanical and geostatistical inferences for mapping natural fracture networks. *Math. Geosci.* 1–27 (2023)
9. Chuginova, T., Corpel, V., Gomez, J.P.: Explicit fracture network modelling: from multiple point statistics to dynamic simulation. *Math. Geosci.* **49**(4), 541–553 (2017)
10. Esteban, A., Zafra, A., Ventura, S.: Data mining in predictive maintenance systems: a taxonomy and systematic review. *Wiley Interdiscip. Rev. Data Mining Knowl. Disc.* **12**(5), e1471 (2022)
11. Grandini, M., Bagli, E., Visani, G.: Metrics for multi-class classification: an overview. *arXiv preprint arXiv:2008.05756* (2020)
12. Hyman, J.D., Karra, S., Makedonska, N., Gable, C.W., Painter, S.L., Viswanathan, H.S.: DFNworks: a discrete fracture network framework for modeling subsurface flow and transport. *Comput. Geosci.* **84**, 10–19 (2015)
13. Jung, A., Fenwick, D.H., Caers, J.: Training image-based scenario modeling of fractured reservoirs for flow uncertainty quantification. *Comput. Geosci.* **17**(6), 1015–1031 (2013). <https://doi.org/10.1007/s10596-013-9372-0>

14. Ke, G., et al.: Lightgbm: A highly efficient gradient boosting decision tree. In: *Advances in Neural Information Processing Systems*, vol. 30 (2017)
15. Liu, X., Srinivasan, S., Wong, D.: Geological characterization of naturally fractured reservoirs using multiple point geostatistics. In: *SPE/DOE Improved Oil Recovery Symposium*. OnePetro (2002)
16. Merembayev, T., Amanbek, Y.: Time-series event prediction for the uranium production wells using machine learning algorithms. In: *56th US Rock Mechanics/Geomechanics Symposium*. OnePetro (2022)
17. Merembayev, T., Bekkarnayev, K., Amanbek, Y.: The identification models of the copper recovery using supervised machine learning algorithms for the geochemical data. In: *55th US Rock Mechanics/Geomechanics Symposium*. OnePetro (2021)
18. Merembayev, T., Yunussov, R., Yedilkhan, A.: Machine learning algorithms for stratigraphy classification on uranium deposits. *Proc. Comput. Sci.* **150**, 46–52 (2019)
19. Mikelic, A., Wheeler, M.F., Wick, T.: A phase-field method for propagating fluid-filled fractures coupled to a surrounding porous medium. *Multiscale Model. Simul.* **13**(1), 367–398 (2015)
20. Nguyen, G., et al.: Machine learning and deep learning frameworks and libraries for large-scale data mining: a survey. *Artif. Intell. Rev.* **52**, 77–124 (2019)
21. Opitz, J., Burst, S.: Macro fl and macro fl. arXiv preprint [arXiv:1911.03347](https://arxiv.org/abs/1911.03347) (2019)
22. Syusyura, B., Box, S.E., Wallis, J.C.: Spatial databases of geological, geophysical, and mineral resource data relevant to sandstone-hosted copper deposits in central Kazakhstan. Technical report, US Geological Survey (2010)
23. Tanha, J., Abdi, Y., Samadi, N., Razzaghi, N., Asadpour, M.: Boosting methods for multi-class imbalanced data classification: an experimental (2020)
24. Wick, T., Singh, G., Wheeler, M.F.: Pressurized-fracture propagation using a phase-field approach coupled to a reservoir simulator. In: *SPE Hydraulic Fracturing Technology Conference*. OnePetro (2014)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

